



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE AMERICAN JOURNAL OF PSYCHOLOGY

Founded by G. STANLEY HALL in 1887

Vol. XXXI

JANUARY, 1920

No. 1

THE LOGIC OF THE NORMAL LAW OF ERROR IN MENTAL MEASUREMENT

By EDWIN G. BORING, Clark University

No amount of practically successful "mental measurement" in laboratories, school-systems, factories or the army can relieve us, if we do not wish to waste time, of the necessity of stopping, every so often, to take account of first principles.

Psychophysics, with more than half a century of history to its credit, has repeatedly found the need to eliminate its logically unfit and reorganize its forces: it is a long cry from the principle of the just noticeable difference to the principle of the psychometric function. The mental test as a newcomer had first to prove its right to consideration. Now that it has been accepted it must pass under the critical eye and learn to conform. And what needs to be said, in way of admonition, applies especially to the mental test, although for no other reason than that the mental test is the lustiest form of mental measurement that one meets today. It, especially, merits a discriminating encouragement.

1. THE NATURE OF PROBABILITY

We can get nowhere in an understanding of mental measurement unless we appreciate the relation of the unit of measurement to the frequency of occurrence of the different measures. In statistics we deal repeatedly with frequency-distributions; and we know, very often, the frequencies before we are assured as to what they are frequencies of. It is not unusual to attempt to calibrate a scale of measurement in terms of frequencies,—an attempt that always demonstrates the

necessity for a definite understanding of the relation between the two. Frequency-distributions, however, have a way of absorbing some of the mystic power that is commonly supposed to inhere in the normal law of error, the 'probability-curve,' which is itself a frequency-distribution. Into the mystery of this function, a mystery connected with its supposed a priori nature, we can pry only by studying its origins, logical and historical. On the logical side we must ask what, in scientific usage, is the nature of a probability. In such a beginning we are setting ourselves no mean task, but a necessary one if we are to avoid muddling.

The early history of the theory of probabilities is a history of the solution of problems arising in games of chance.¹ An applied mathematics of probability preceded the pure. For us the most significant principle that this period brought forth is one that is implicit in the famous theorem of James Bernoulli (1713).

This theorem holds that events tend to occur with relative frequencies proportional to their probabilities. If a is twice as likely as b to happen, then in the long run it will happen twice as often. But how, we may well ask, are we to know the probabilities of occurrence for a and b ? We must exclude the inverse use of the Bernoullian theorem. We can not say that a is twice as likely to happen as b because it happens twice as often, for this is simply to reason from frequency to probability and ultimately to identify the two. Our problem was to predict frequency from given probability. How then do we achieve this given probability?

The case of the coin appears at first glance fairly obvious.

If we may regard an ideal coin as a uniform, homogeneous circular disc, there is nothing which can tend to make it fall more often on the one side than on the other; we may expect, therefore, that in any long run of throws the coin will fall with either face uppermost an approximately equal number of times, or with, say, heads uppermost approximately half the times.²

Thus we read in a modern textbook of statistics. The thesis is: the probabilities are equal when there is no good reason for their being unequal. Von Kries has called this law the principle of insufficient reason (*"das Princip des man-*

¹ I. Todhunter, *History of the mathematical theory of probability*, 1865, is the standard historical work. He begins with Pascal and Fermat in 1654 and continues to Laplace: 624 pp.

² G. U. Yule, *Introduction to the theory of statistics*, 4th ed., 1917, 256.

gelnden Grundes");³ Boole speaks of it as "the equal distribution of ignorance."⁴

Against an indiscriminate use of the principle of insufficient reason von Kries inveighs. We are asked to consider before the days of spectroscopic analysis the question of whether there is iron in Sirius. We are completely ignorant, but we can conceive of no greater reason for the presence of iron than against it. The probability for the existence of iron in Sirius is therefore $\frac{1}{2}$, on the principle of insufficient reason. Similarly the probability for the existence of gold is $\frac{1}{2}$. It follows algebraically that the probability for the existence of both iron and gold in Sirius is $\frac{1}{4}$, of iron without gold $\frac{1}{4}$, of gold without iron $\frac{1}{4}$, and of neither of these metals $\frac{1}{4}$. There were, when von Kries wrote, 68 known earthly elements. The probability of the existence of all these elements in Sirius is the 68th power of $\frac{1}{2}$; *i. e.* the chance of all the elements existing in Sirius, as well as the chance of no one of them existing there, is about three in one-hundred-billion-billions. But if we had raised in the first place the question of the existence of earthly elements in Sirius, we might, being ignorant and applying the principle of insufficient reason, have said that it is just as likely that there are no earthly elements in Sirius as that there are some. The discrepancy between the three one-hundred-billion-billionths and one-half leads von Kries to declare that "die Aufstellung der gleich möglichen Fälle muss eine in zwingender Weise und ohne jede Willkür sich ergebende sein."⁵

Plainly it must. Shutting our eyes reduces our faith in what we see to less than one-half. If we would know the nature of our surroundings, we must open our eyes; and, if we would acquire a tenable belief of the chemical constitution of Sirius, we must await the spectroscope. There is no alchemy of probabilities that will change ignorance into knowledge. Expectations must be founded upon cogent rather than insufficient reasons. Nevertheless we can not, it would seem, push this antithesis too far. We can never be completely informed of all the conditions. Thus Westergaard:⁶

³ Joh. von Kries, *Die Principien der Wahrscheinlichkeitsrechnung*, 1886, 6. The book is "eine logische Untersuchung." See also A. Kaufmann, *Theorie und Methoden der Statistik*, 1913, 43ff. C. Stumpf, in *Sitzungsber. d. k. bayr. Akad. zu München (philos.-philol. Cl.)*, 1892, 37-120, in arguing against von Kries, appears to defend the principle of insufficient reason; but in so doing he appeals to a wider analysis of the situation that is not wholly unlike a principle of cogent reason. See especially 41, 61-79.

⁴ G. Boole, *An investigation of the laws of thought*, 1854, 370.

⁵ *Op. cit.*, 10f.

⁶ H. Westergaard, *Grundzüge der Theorie der Statistik*, 1890, 4.

In every well-arranged game of chance . . . the balls are of the same size, of the same wood, of the same specific gravity, *etc.* They are mixed with the greatest care and every ball is apparently subject to the same forces. Yet such is not the case. In spite of all effort the balls are different, and depart, even though quite insignificantly, from spherical form. One approaches this and another that mathematical form. Every one has a weight and size different from all the others. No ball is absolutely like any of the others. Moreover they can not possibly lie in the bag in exactly the same manner. In short there are a multitude of apparently insignificant differences which determine that exactly this ball and none other shall be drawn.

In the face of such complexity it is hopeless to seek cogent reason for the drawing of the particular ball.

Moreover, there is a logical as well as a practical reason for our inability to predict in the individual case. Westergaard seems to see this when he implies—it amounts to this—duplicity on our part when we arrange the game of chance. We attempt to make the balls all equal so as to equalize chance, we assume that we have succeeded for all practical intents and purposes, and yet in the same breath we expect to be able to draw some particular one of the balls. However, there must be inequalities, including of course difference of spatial position within the bag, in order to make the drawing possible.

Did no such inequalities exist, then one of two things would happen: either all the balls would turn up at once or they would all remain in the bag.⁷

In more general terms, we may say that the problem of probability exists only in the face of ignorance. Given all the conditions, we deal not in probabilities but with the certainties of cause and effect.⁸

Here then it would seem we are landed in a dilemma. If we are in total ignorance of the conditions of our problem, all argument is foolishness. If we know all the conditions, we may argue to an effect but not to a probability. Have we to deal then either with a practical impossibility or with a logical absurdity?

The solution lies in the adoption of a definite intermediate position. Fisher lays down a first principle in the case:

It thus appears that a rigorous application of the principle of cogent reason seems impossible. However, a compromise between this principle and that of the principle of insufficient reason may be effected by the following definition of equally possible cases, viz.: *Equally possible cases are such cases in which we, after an exhaustive analysis of the physical laws underlying the structure of the complex of causes*

⁷ *Loc. cit.*

⁸ Thus J. Venn, *Logic of chance*, 3rd ed., 1888, discusses probability as "amount of belief," where "ignorance of the individual is presupposed" and "causation . . . denied, within considerable limits:" p. 122.

*influencing the special event, are led to assume that no particular case will occur in preference to any other.*⁹

But Fisher, in attempting a compromise, neglects to tell us definitely how much ignorance and how much knowledge make the proper conditions from which we may proceed.¹⁰ We are scientists engaged in very practical work and we wish to know what we may and what we may not do in our statistics.¹¹

We can get at the nature of probability by taking a simple instance. The event in which we are interested is, let us say, the toss of a penny. We wish to know the probability that it will come heads. The law of insufficient reason says that, since we can not know that the penny is more likely or less likely to come heads than tails, the probability of heads is $\frac{1}{2}$. The extreme demand for cogent reason would urge that, since we can not know the principal determining factors, this case is insoluble. The 'compromise' that Fisher proposes is that we shall analyze the underlying physical laws. We shall find that one side of the penny is heads and the other tails, that the penny is homogeneous, that it spins many times in the toss, and so on, until we are "led to assume" that neither heads nor tails will be preferred. Now the crux of the whole matter lies in the mechanism by which we are led to this assumption. In the case of the penny no knowledge of physical constitution and of the complexity of tossing would ever be valid ground, in itself, for believing that the penny would come heads as often as tails. In the initial instance this fact

⁹ A. Fisher, *Mathematical theory of probabilities*, 1917, vol. I, 9. An interesting book, written with historical and philosophical perspective for practical purposes. It emphasizes the work of Continental statisticians in a way that English writers do not. Volume II seems not yet to be published.

The position is approximately Kaufmann's, *loc. cit.*, in distinguishing between *mangelnder* and *zwingender Grund*.

¹⁰ It will not keep the statistician from error to tell him that his analysis must be "exhaustive" or that it must be so exhaustive as to 'lead him to assume . . .' Not only do investigators differ in natural thoroughness, but the problem may be restated, as we shall see, so that analysis of the properly known conditions may be pushed to the limits of cogency.

¹¹ The present paper in its inception is an attempt to answer certain questions that arose in the Office of the Surgeon General of the U. S. Army with respect to the ratings of recruits on intelligence tests. Actually logic is not a far cry from practice, but the technologist upon occasion can be deaf.

must be observed.¹² Thereafter we may reason that other homogeneous, one-headed, well-tossed pennies will in the long run, come heads fifty per cent of the time, and we may even go farther and reason that a homogeneous, six-faced, well-tossed die will in the long run come up equally often on every face. What we do is to assume probability on the grounds of observed frequency and its conditions.

Let us state the case in roughly syllogistic form.

The tossing of a penny gives, when indefinitely repeated, 50% heads;
The event in which we are interested is a tossing of a penny;

The event in which we are interested gives, when indefinitely repeated, 50% heads.

This leads us to predict what will happen to our particular penny, if it is tossed repeatedly; and, what is more important, it defines for us the place of cogent reason and the sphere of ignorance. We must know what a penny does when tossed again and again, either directly from observation or indirectly by inference from other observations. The major premise must be based upon cogent reason and there is no limit to the exactitude of the knowledge. Ideally the major premise may be considered to be indubitably established. Moreover we must know that the event in which we are interested is the toss of a penny in the sense of that term in the major premise. Our minor premise, by "an exhaustive analysis of the physical laws underlying the structure of the complex of causes influencing the special event," must also be established and without limit of accuracy by cogent reason. Of what then do we remain ignorant? Of the manner in which the particular, in the middle term, is subsumed under the general. In other words we know that our event is such an event that if repeated it will yield in the long run heads and tails equally, but we do not know whether this particular event, unrepeated, will be heads or tails.

Of an equal distribution of this ignorance, of a law of insufficient reason, we are saying nothing. But have we attained to a statement of probability by saying of a single event that

¹² "The assumption that any probability-constant about which we know nothing in particular is as likely to have one value as another is grounded upon the rough but solid experience that such constants do, as a matter of fact, as often have one value as another." F. Y. Edgeworth, *Mind* O. S. 9, 1884, 230. K. Pearson approves this view: *Grammar of science*, 3rd ed., 1911, pt. i, 146. See also Edgeworth in article on Probability, *Ency. Britannica*, 1911, xxii, 377, 391. There have been many attempts to supply "the rough but solid experience" empirically; e. g. Edgeworth, *Cam. Philos. Trans.* 20, 1905, 128ff.; Pearson, *Chances of death*, 1897, i, 13ff., 48; Westergaard, *op. cit.*, 21-56.

it is such that, if repeated, it would yield a given frequency of particular cases?

We can not within the narrow limits of this paper escape a certain amount of dogmatism. We must read widely elsewhere on the logic of chance. This much may, however, be said. It is plain that the probability of the event can not inhere in the event. The coming throw is either heads or tails, not half heads and half tails. An event is not even conditioned upon its probability. If, from a box of one thousand tickets, numbered consecutively, one draws any ticket at all, one obtains that particular ticket in the face of odds of 999 to one against it. The rule of such a drawing is that the improbable always happens. If the probability does not lie in the event, it must lie outside it; and we may here assert that it lies in the series to which the event belongs. When we ascribe a probability to a particular event we are simply seeing that event in a series in which the event is repeated as it varies in some particular phase; or, to put it more picturesquely, we see a series of repeated events telescoped within the single instance. This is our habitual mode of dealing with ignorance of this kind. Unable to prophesy the particular event, but able to prophesy its frequency, we content ourselves with the general in view of our impotence to deal with the particular.¹³

(And we may note in passing that ignorance has not bred knowledge in this process. Where reason was insufficient, prediction failed. We never can tell what will turn up next. We may be able to tell what will happen one way often and another way seldom, but that tells us not at all what the next one is).

¹³ The paragraph is compatible with Venn's chapter on *Measurement of belief*, *op. cit.*, 119ff., *q. v.*, although divergent in emphasis. See also the discussion by F. M. Urban, *Ueber den Begriff der mathematischen Wahrscheinlichkeit*, *Vierteljahrschrift f. wiss. Philos. u. Soziol.* 35, 1911, 1-49, 145-185. Only the margins of the logical problem are skirted here. Much might be said of subjective and objective probability. For the scientist-psychologist it is perhaps useful to distinguish four kinds of probability: (1) observed frequencies; (2) expected frequencies; (3) participation in a series of expected frequencies; and (4) psychological expectation. (1) is scientific fact; (2), derived from (1), is a prediction; (3) is a probability, a 'telescoped' (2); (4), the actually observable mental expectation, is a function of (3) and other conditions for determining tendencies, which the psychologist might do well to investigate. See Venn, 152-162. The metaphysical problem of the relation of physical constitution to frequency is beyond the point; actually the relation is observed. In more exact usage it would be necessary to assume that the phrase, 'indefinite repetition,' be taken to mean that the given frequency is approached as a limit when the repetitions are multiplied. The approach would not be continuous; Venn, 118.

We may seem far from the subject of mental measurement, but not without occasion, as we shall see. Let us turn now to the part played by the normal curve of error.

2. THE ROLE OF THE LAW OF ERROR

The normal law of error¹⁴ has been both an inspiration and a limitation in statistical measurement. Its formulation by Laplace resulted during the last century in a wide extension and application of the law to numerous forms of scientific, social and biological measurement; but Laplace is not responsible for all that occurred. There is a bit of magic in the formula. The law came to play the part of a first principle of nature, of an ideal, given *a priori*, to which nature seeks to conform. The mathematicians wrought slowly, but they wrought a god. Against such blind faith later statisticians have protested. They call the normal law a "fetish" and its *a priori* use a "superstition." Nevertheless the "superstition" still lingers and is mixed up with mental measurement. For this reason we are going to enquire, concerning the law of error, what real value it has for us to-day as a scientific tool.

In the days when the application of the theory of probabilities was largely confined to games of chance,¹⁵ the logical difficulties discussed above had little opportunity to make trouble. The fact is that the cards in a pack, the faces of a die, and the sides of a coin do in the long run and under the conditions of gaming turn up about equally often. At this level common sense takes care of the theorist. It makes no practical difference whether the reason he assigns for the equality of occurrence is his lack of sufficient reason for any other result or the empirical fact that these things do work in such a way. Nor does it matter if unwittingly he applies to the single event the frequency that belongs to the series and calls it the probability of the event, for he knows by common sense that probability is not prophecy and that he

¹⁴ The so-called "Gaussian" curve. The mathematical propaedeutics for this function were prepared as long ago as the beginning of the 18th century (De Moivre, 1718). See Todhunter, *op. cit.* 71f., 136, 191ff., 552f. Laplace's *Theorie analytique des probabilités* brought together and supplemented in 1812 his work of thirty-five years, and gives the development of the law of error on pages 275ff. The proper date for the law is presumably 1786; see Todhunter, 485. Gauss gives the formula in Article 178 of the *Theoria motus corporum coelestium*, 1809, for which see his *Abhandlungen zur Methode der kleinsten Quadrate*, 1887, 102f. On his ascription of credit to Laplace, see *ibid.*, 207.

¹⁵ The period covered by Todhunter, *op. cit.* Todhunter leaves off unfortunately where our most important period begins.

can not be assured of a realization of his expectations except in the long run. The fact is that the principle of insufficient reason and an indefinite conception of the nature of probability do when applied to games and modified by such unconscious reservations, furnish safe ground for practice, even though the logical implications are dubious.

Laplace and Gauss extended the theory of chance to the probability of errors.¹⁶ Errors are a new material and it is not plain at once that the old assumptions hold. So far as the development of the law of error goes, these mathematicians were content to make their assumptions and reason abstractly to the general conclusion. Laplace worked in the tradition of the old school. He developed the law of error as the limiting case of the binomial expansion when the number of terms is infinite. This process is equivalent to assuming that, in a given case, an error is the resultant of a large number of sources of error, every one of which may affect the result in one way or the opposite (or affect it or not affect it), *i. e.* every source of error may, in the particular case, come up "heads" or "tails," as it were.¹⁷ Either we have, in practice, to assume the principle of insufficient reason, which, applied to errors, does not have the sanction of common-sense experience; or we have to achieve an analysis of the case into sources of error and determine for every one that its law is the law of the coin, a course which is ordinarily impossible. Gauss hypothesizes explicitly that positive and negative errors are equally likely, and derives the law on this basis. Logically his assumption is the same as Laplace's, but it is usually stated in a form that suggests a test. If positive and negative errors are equally likely, then the arithmetical mean is the most probable value. To test the validity of the assumption in practice, we have then only to select a series (of observations, say, which are subject to error) and see whether the arithmetical mean occurs most frequently.

Laplace's derivation suggests the nature of error, Gauss' its mode of manifestation. Both these famous mathematicians were arguing from premise to conclusion when they derived the law of error, and science remains in their debt. We have no complaint to make until we find the law applied without an effort to establish the premises in the particular case. *The*

¹⁶ Cf. W. S. Jevons, *Principles of science*, 1883, 375-385.

¹⁷ Jevons, *op. cit.*, 380f. L. D. Weld, *Theory of errors and least squares*, 1916, 41-56, has an elementary discussion of the binomial expansion and the law of error. For the logic of this derivation applied to psychophysics, see E. G. Boring, *Am. J. Psychol.* 28, 1917, 465ff.

law will not work for errors, unless errors fulfill its conditions.

Gauss was clear about his assumptions, but he did not hesitate to apply the law to astronomical errors of observation.¹⁸ Some of his followers, it seems, who had not witnessed the birth of the law, were less concerned in its application with its parentage. In theory of measurement the law was often accepted on its face value, in spite of the fact that a constantly increasing list of exceptions to it was being made out. In such cases the argument for the law seems to have been the practical one that it gave but little trouble in actual use. Thus Urban recently wrote:¹⁹

We are confronted by the shocking situation that a proposition is triumphantly borne out by an immense indirect experience and that it can be proved neither by mathematical deduction nor by direct experience.

Urban, however, did attempt to prove it by showing, under the conditions of the psychophysical experiment, that positive and negative errors of observation, measured from the point of subjective equality, are equal.²⁰ Faith in the principle was established, however, long before psychophysics furnished this justification.

It is upon Quetelet, it seems, that we must fix the responsibility for the uncritical extension of the normal law to various human measures other than errors.²¹ In making a man, he holds, Nature aims at an ideal, and the differences

¹⁸ On the assumptions: Gauss, *op. cit.*, 4; on the applications: 92ff., 129, 139ff., etc.

¹⁹ Urban, *Psychol. Rev.* 17, 1910, 242.

²⁰ *Op. cit.*, 240-244.

²¹ An interest in the application of theory of probabilities to a variety of human problems was of long standing. Todhunter mentions the following initial attempts at application: expectation of life (Graunt, 1662), human testimony (Craig, 1699), human innocence (N. Bernoulli, 1709), birth-rates (Arbuthot, 1710-12), sex-ratios (De Moivre, 1718), astronomical observation (Simpson, 1757), marriages, small-pox, and inoculation (D. Bernoulli, 1760), weather forecasts (Lambert, 1773), annuities (Lagrange, 1798), modes of election (Laplace, 1812), and scientific observation in general (Laplace and Gauss). The important names are N. Bernoulli, De Moivre, D. Bernoulli, Lagrange, Trembley, and Laplace. See Todhunter, *op. cit.*, 37, 41, 54f., 193, 196, 211, 228f., 265, 320, 335, 349, 423, 426, 446, 500f., 546, 589-613, 615ff.

Laplace in the *Essai philosophique sur les probabilités*, which is a separately published introduction to the *Théorie* (*op. cit.*), devotes space to the application of probabilities to natural philosophy, to the moral sciences, to testimony, to elections in assemblies, to the judgments of tribunals, to tables of mortality, and to the duration of life, marriages, etc. Gauss is interested in applications to astronomy and geodesy. But all these things were more or less subject to a law of errors. Quetelet's conception of any natural variation as an error is

between men represent her degrees of error.²² Quetelet refers for evidence to the measurement of circumferences of the chests of 5,738 Scotch soldiers and of the heights of 100,000 French conscripts. His position would be all very well were Quetelet content with supporting his hypothesis by this evidence. He seeks, however, to make it work both ways. He suggests that the deviation of the curve for the French conscripts from the normal law is due to fraudulent rejection for military service.

This . . . law of continuity enables us to recognize a more remarkable fact: we might suspect it, but here we find it proved,—it is that the number of men rejected for deficiency in height is much exaggerated. Not only can we prove this, but we can determine the extent of the fraud. The official documents would make it appear that, of 100,000 men, 28,620 are of less height than 5 feet 2 inches: calculation gives only 26,345. Is it not a fair presumption, that the 2,275 men who constitute the difference of these numbers have been fraudulently rejected?²³

Of course it is not a fair presumption, since the law depends on the official documents for confirmation. Nevertheless, while admitting the dependence of the law on experience, Quetelet proceeds in numerous cases to analyze experience by means of it. Such a double-edged sword is a peculiarly effective weapon, and it is no wonder that subsequent investigators were tempted to use it in spite of the necessary rules of scientific warfare.

Galton in the *Hereditary Genius* applied the normal law to

an undoubted extension of the principle, although it is difficult to draw a line.

A. Quetelet was a Belgian statistician, who published an *Essai de physique sociale* in 1835 and a collection of *Lettres sur la théorie des probabilités* in 1846. He writes in the tradition of Laplace with little reference to Gauss. That Gauss' work on least squares was of interest to the French school is evidenced by its translation into French (from the Latin) in 1855, thirty-two years before a similar publication in the German language.

There is no question as to Quetelet's responsibility for the extension of the principle: cf., e. g., W. Lexis, *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*, 1877, 11; Kaufmann, *op. cit.*, 12, 163.

²² *Letters on the theory of probabilities*, trans. 1849, 91-96. "The difference which Nature makes in the heights of men is not greater than that which inexperience would produce in the measurements taken on one individual man in an attitude more or less curved." "Everything occurs as if there existed a type of man from which all other men differed more or less." "Each people presents its mean, and the different variations from this mean, in numbers which may be calculated *a priori*." (pp. 95f.)

Cf. the discussion and criticism of "l'homme moyenne" in Kaufmann, *op. cit.*, 110ff.

²³ *Letters*, pp. 97f.

mental differences and, using it *a priori*, worked from frequencies of natural ability to a scale of equal intervals of ability. He frankly accepted "the very curious law of 'deviation from an average'" from Quetelet, showed that certain mental capacities obeyed the law approximately, and then argued from the law to "grades of natural ability, separated by equal intervals."²⁴

The English school of biometricians, which derives directly from Galton, has furthered the *a priori* use of the normal law that obtains to-day in much biometric and mental measurement.²⁵ There is an early study by Pearson that illustrates the point.²⁶ Weldon had measured the lengths and breadths of two large samplings of crabs: one set from Naples and one from Plymouth. The Plymouth batch gave symmetrical distributions, but the frequency curve for the breadths of the Naples batch was skewed. Pearson, by an elaborate mathematical treatment, analysed the skew curve into the sum of two normal curves, the averages of which did not coincide, and Weldon accepted this result as evidence of "dimorphism." Pearson then applied his method to a symmetrical curve for the Plymouth crabs and found that there was no real solution. He assumed, therefore, that these data were "homogeneous." Further he laid down the general rule that some such analysis must be attempted upon every apparently normal curve before a conclusion can be reached as to the homogeneity of the data. If the actual results can be better represented by the sum of two normal curves than by any single normal curve, then there is heterogeneity. Even curves that approximate the normal very closely must be tested, for they may yield to analysis.

Thus Pearson showed that the distribution of shots at a

²⁴ F. Galton, *Hereditary Genius*, 1869, 26-36. G. Th. Fechner, of course, had already used the law in the method of right and wrong cases with sensory material. He seems not to have hesitated in his acceptance of the law from Gauss: *Elemente der Psychophysik*, i, 1859, 104. He applied it, not to sensations, but to the *scheinbare Reizgrösse*, the effective inner stimulus: *Revision der Hauptpunkte der Psychophysik*, 1882, 40f., 45. Current notions of "deviation from an average" come undoubtedly, however, from Galton and Quetelet.

²⁵ See, for example, M. R. Trabue, *Completion-test language scales*, 1916, 29-60, where the normal law is assumed in order to obtain a scale of equal units. Trabue is quite explicit about the assumption (see note, p. 31), but he evidently believes it is correct or he would not devote so much space to developing it.

²⁶ K. Pearson, Contributions to the mathematical theory of evolution, *Philos. Trans.* 185A, i, 1894, 71-110. The basal data are given by W. F. R. Weldon, *Proc. Roy. Soc.* 54, 1893, 318-329.

target, given by Merriman,²⁷ is best represented by the sum of two normal curves. The gun must be thought of as aiming, not at the sixth band of the target as Merriman supposed, but at the fifth and seventh bands (explainable perhaps by a change of sighting during mid-firing). Here then, we find no attempt at all to seek empirical support for the normal law. The crabs were obtained in a single selection and the shots were made in a single experiment, and each set appears, even after it is charted in a graph, as if all the measures belonged together. But we apply the criterion of the normal law and we find the crabs dimorphic and the gun double-sighted.²⁸

Now the very remarkable thing about this easy acceptance of the law of error as the rule of nature is that it has been done all along in spite of very much better available information. Let us get some idea of what has been said.

We saw that Laplace's assumptions were implicit in his method and Gauss' explicit in his text. We had no quarrel with them, but we blamed Quetelet. Did Quetelet's thesis pass, then, quite unchallenged? No, for in the notes to the *Letters* he printed three letters from Bravais,²⁹ who presented figures on the skew distribution of barometric pressures, discussed various reasons for departures from the normal law, and said in brief:

I think that generally every partial and distinct cause of error gives place to a curve of possibility of errors (or, if preferred, of differences about the mean), which may have any form whatever,—a curve which we may either be able or unable to discover, and which in the first case may be determined by consideration *a priori* on the peculiar nature of this cause, or may be determined *a posteriori* by observations on the isolated condition of other concomitant causes of error.³⁰

²⁷ M. Merriman, *Textbook of the method of least squares*, 1884, 14. These data have frequently been quoted as illustrating the application of the normal law to errors.

²⁸ Of course, if we wish to define "heterogeneity" solely by reference to the applicability of the normal law, nothing can be said, though the usefulness of the procedure might be questioned. The almost unavoidable implication of such a course, however, is exactly that which Weldon got from it, *viz.* that the two normal components point to a biologically significant dimorphism. That idea, moreover, originated with Quetelet: *Letters*, 96f. On the other hand, it is only fair to Pearson to point out that the theory of skew variation in homogeneous material must already have been taking shape; see his distinction in the year following between asymmetry due to heterogeneity and asymmetry due to the mode of variation of a homogeneous material: *Philos. Trans.* 1862, i, 1895, 344f.

²⁹ Quetelet, *Letters*, 286-295. Venn, *Nature*, 36, 1887, 411f., shows the same asymmetry of barometric heights that Bravais pointed out.

³⁰ P. 290.

In introducing the letters from Bravais, Quetelet remarked:

In this memoir I only had in view the examination of the case in which accidental causes have had no tendency to act in one direction more than another.³¹

We have already shown that Quetelet actually established the law by the data and then corrected the data by the law.

In 1869 Galton accepted the law of deviation from an average as applicable to mental ability, but in 1879 he was ready to insist on exceptions:³²

My purpose is to show that the assumption which lies at the basis of the well known law of "Frequency of Error" . . . is incorrect in many groups of vital and social phenomena, although that law has been applied to them by statisticians with partial success and corresponding convenience. . . .

The assumption to which I refer is, that errors in excess or deficiency of the truth are equally probable; or conversely, that if two fallible measurements have been made of the same object, their arithmetical mean is more likely to be the true measurement than any other quantity that can be named. . . .

Suppose we endeavor to match a tint: Fechner's law, in its approximative and simplest form of sensation= \log stimulus, tells us that a series of tints, in which the quantities of white scattered on a black ground are as 1, 2, 4, 8, 16, 32, &c., will appear to the eye to be separated by equal intervals of tint. Therefore, in matching a grey that contains 8 portions of white, we are just as likely to err by selecting one that has 16 portions as one that has 4 portions. In the first case there would be an error in excess, of 8; in the second there would be an error in deficiency, of 4. Therefore, an error of the same magnitude in excess or in deficiency is not equally probable.

Thus Galton argues that the geometric mean is the most probable value with sensory material or data that follow a similar law. The distribution curve in such cases would be markedly skew.

We remarked that Pearson in 1894 gave evidence of being influenced by the sanctity of the normal law. His faith was not great for he noted its limitations in 1895.³³ In 1900 the faith was gone.³⁴

It appears to me that, if the earlier writers on probability had not proceeded so entirely from the mathematical standpoint, but had endeavored first to classify experience in deviations from the average, and then to obtain some measure of the actual goodness of fit provided by the normal curve, that curve would never have obtained its present position in the theory of errors. Even today there are those who regard it as a sort of fetish; and while admitting it to be at fault

³¹ P. 286.

³² *Proc. Roy. Soc.* 29, 1879, 365-367. The paper is followed by a discussion by D. McAlister, *ibid.*, 367-376 which gives the mathematics and presents some graphs of skew distributions that would result from the logarithmic scale of measurement.

³³ *Philos. Trans.*, *loc. cit.*

³⁴ *Philos. Mag.*, 5th ser., 50, 1900, 173.

as a means of generally describing the distribution of variation of a quantity x from its mean, assert that there must be some unknown quantity z of which x is an unknown function, and that z really obeys the normal law! This might be reasonable if there were but a few exceptions to this normal law of error; but the difficulty is to find even the few variables which obey it, and these few are not usually cited as illustrations by the writers on the subject.³⁵

Pearson in developing a system of skew curves whereby he seeks to adjust theory to nature has performed great scientific service in combatting the superstition of the normal law.³⁶ The paper of 1894 suffered on account of its occasion, *viz.*, the problem of dimorphism in Weldon's crabs. But not all investigators have avoided initial error; and presently we must return again to Pearson.

These quotations from Galton and Pearson suggest a reason why the normal curve should be of exceptional occurrence. If "log stimulus" gives a normal curve, says Galton, then "stimulus" will not. If x does not give a normal curve, some persons, says Pearson, assert that z , of which x is a function, does follow the law. Suppose then that nature does conform to the normal law; we shall not in practice be fortunate enough to obtain normal curves of distribution unless we have chanced to use nature's own unit of measurement, or some unit directly proportional to it. Bertrand pointed this out long ago.³⁷

La règle des moyennes, il importe d'insister sur ce point, n'est ni démontrée ni exacte. S'il était admis que la moyenne entre plusieurs mesures fût toujours la valeur la plus probable, il en résulterait des contradictions. Quand on mesure une grandeur, on mesure, par cela même, toutes les fonctions de cette grandeur, son carré par exemple, ou le logarithme du nombre qui la représente. Pourquoi la valeur la plus probable du carré ne serait-elle pas la moyenne des valeurs obtenues pour le carré, et la valeur probable du logarithme, la moyenne des logarithmes? . . .

Il ne faut pas, pour écarter l'objection, faire une distinction entre les grandeurs directement mesurées et celles qui résultent d'un calcul. Un mécanicien pourrait, bien aisément, annexer à une balance une aiguille marquant le carré ou le logarithme du poids. Ce carré ou ce logarithme deviendrait alors la grandeur mesurée. Le postulatum admis dans un cas devient donc impossible dans les autres.

A better example would be the distribution in size of the cubical crystals of common salt. In a sample of such crystals we might measure size by the height of the cube or by its

³⁵ The remarks are occasioned by Merriman and the target-shooting. Merriman has subsumed these data under the "fetish"; but we have seen that it is not perfectly clear why Pearson in 1894 subsumed the same data by a two-fold application of the same "fetish."

³⁶ The early papers on "Skew variation in homogeneous material" are *Philos. Trans.* 186A, i, 1895, 343-414, and 197A, 1901, 443-459. See also "On the general theory of skew correlation and non-linear regression," *Drapers' Co. Research Memoir*, 1905.

³⁷ J. Bertrand, *Calcul des probabilités*, 1889, 180f.

weight. Since the latter measure should be approximately proportional to the cube of the former, it is plain that the normal law could not hold in both cases.³⁸ Now, if nature seeks to conform to the normal law, does she prefer to conform in lengths or in volumes or, perhaps, in some other more "natural" unit? At least, if man would seek the law of nature, he must have cogent reason for his choice of measures.³⁹ He can not (as the behaviorist would agree!) tell what nature is doing unless he knows what nature is trying to do.

We have been speaking of France and England, but the reaction against the "Gaussian dogma" was more explicit in Germany. In 1877 Lexis published a *Theorie der Massenerscheinungen*⁴⁰ in which frequency distributions were dealt with as normal, subnormal, or supernormal.⁴¹ The emphasis was thus shifted from the normal distribution to normal deviations from the normal distribution.⁴² The "Lexian ratio" measures this deviation.⁴³ This point of view has met with acceptance among German and Scandinavian actuarial statisticians, and forms, for example, the basis of Czuber's *Wahrscheinlichkeitsrechnung*.⁴⁴ We have already seen that von Kries in 1886 did not include among valid logical principles the law of insufficient reason. He was inclined to believe, however, that in many cases the "Gaussian law" was realized,⁴⁵ and he cites in support Bessel's observations, which Urban refers to as an example of asymmetry.⁴⁶

Westergaard in 1890 sums up the empirical case.⁴⁷ He

³⁸ It would be interesting to test this relation in the case of heights and weights of men. Height-and-weight tables are available in anthropometrical and somatological texts.

³⁹ We have a cogent reason, presumably, when, with the Fechner pendulum, we take as the unit, not the angle read on the scale, but the square of the sine of half that angle.

⁴⁰ Lexis, *op. cit.*; see especially 34ff.

⁴¹ Fisher's translation of *unternormal* and *übernormal* as "subnormal" and "hypernormal" is due presumably to his otherwise excellent "dano-English," for which he so modestly apologizes: *op. cit.*, 125 and xv.

⁴² If a pun is permissible, for Lexis' thesis is that the "normal curve" is not "normal" in the sense "usual."

⁴³ V. Fisher, *op. cit.*, 124ff.

⁴⁴ E. Czuber, *Wahrscheinlichkeitsrechnung*, 2 Aufl., 1910, 34-78.

⁴⁵ *Op. cit.*, 226ff.

⁴⁶ *Op. cit.*, 241. Much of the dispute as to whether the normal law is realized in fact, depends on just how much departure of fact from theory we allow without giving up the theory. Von Kries could call normal in 1886 what Urban would see as asymmetrical in 1910: standards have gotten more rigorous.

⁴⁷ *Op. cit.* 10-83.

brings together the experiments on the drawing of balls and the like, and shows that with increasing number of cases the relative frequencies lie within a decreasing *Spielraum* and thus in a sense approach the expected frequencies. He then goes on to show that the frequencies of social statistics behave in a manner similar to the frequencies of the experiments on "chance." He is not so much interested in the applicability of the normal law as he is in the facts. The appeal to facts is, however, always a protest against theory which is given *a priori*.

After all the facts are the scientific business. We have traced what seemed to be a growing realization of the logical inadequacies of a theory, but have we perhaps not been witnessing a deepening appreciation of the facts? When anthropometric measurements were first being made, there stood out the fundamental fact of the massing of cases about the average, the rapid falling off of more extreme frequencies, and the extreme rarity of widely divergent cases. To this extent the normal law was the fact. But as interest centered upon the details, the inadequacies of so simple a generalization became apparent. There was effort enough at reconciliation, but, in general, science kept to the facts, and a more flexible system of representations came into use.

There is a pretty instance of the inadequacy of the normal law to scientific description in a research of Pearl's.⁴⁸ Pearl studied statistically the number of leaves in a whorl in *Ceratophyllum*. He found that the distributions for number of leaves were skewed one way at the proximal end of the main stem or of a branch, the other way at the distal end, and that the form of the distribution passed through symmetry somewhere between the two ends. Thus he writes:

The phenomenon of skew variation stands forth in this case, free of doubtful interpretation through selection or any similar factor, clearly and definitely as a phenomenon of growth. In the face of facts of this kind it is difficult to understand how anyone can be so firmly convinced of the *Allgemeingültigkeit* of the normal or Gaussian law, as some biologists still are Skewness in variation is a very real biological phenomenon, which may be changed and modified, not only in degree, but in direction, by various biological factors like growth, as, for example, in the present case.

In other words, if nature is aiming to make something follow the normal law, then, as the stem of *Ceratophyllum* grows, she changes continuously that something which is to fit the law. The real problem of scientific description is an account

⁴⁸ R. Pearl, *Variation and differentiation in Ceratophyllum*, 1907; especially 90f.

of nature's changing. To assume the normal law in such a case would be to shut one's eyes to the fact.

Here we may leave the question of the *a priori* nature of the normal law. There is, after all, no magic in it. It gives us back always what we put into it. If we know from experience what nature is up to, as we do with the coin, then we can proceed upon cogent reasons to apply the law and we get results. If we do not know, we must appeal to nature and see. We have then no reason to expect that we are going to find the law before the appeal is made, whereas we have considerable reason to expect not to find it, since the form of distribution depends on the unit of measurement and we have arbitrarily chosen one of a possible infinite number of units. When we do go to nature we find all degrees of resemblance to the law and divergence from it, and we may even find that the degree of divergence from the normal becomes the significant fact of our observation.⁴⁹

3. THE RELATION BETWEEN THE UNIT OF MEASUREMENT AND THE FORM OF DISTRIBUTION

We have seen that if we alter the unit of measurement we alter the form of the frequency-distribution, and that if the change is in accordance with any function other than direct proportionality we get alterations of the skewness of the curve. Bertrand made the point with especial clearness, and it is strange that a relation so obvious should have received so little attention.

Williams recently ran upon the difficulty when experimentally he obtained psychometric functions for memory.⁵⁰ He

⁴⁹ We have noted that the universality of the normal law was contested by the following persons: Bravais, 1845; Quetelet (admitted Bravais' argument), 1846; Lexis, 1877; Galton, 1879; McAlister, 1879; Bertrand, 1889; Pearson, 1895 and emphatically in 1900; Pearl, 1904; Urban, 1910; Edgeworth, limitedly in the *Britannica*. The list is by no means complete. We may add: R. L. Ellis, *Philos Mag.*, 3 ser., 1850, 321-328; Jevons, *op. cit.*, 1883, 374ff.; Venn, *Nature*, 36, 1887, 411ff. and *Logic of chance*, 1888, 23-52, 501ff.; Fisher, *op. cit.*, 1917, 149. Urban, *op. cit.*, 241, mentions Helmert, Laurent, Bessel, Guarducci, and Edgeworth as demonstrating asymmetry of distribution. Fisher, *loc. cit.*, refers to Oppermann, Cournot, and Dormay (especially Dormay) as showing the inadequacy of the Gaussian dogma. Urban's serious consideration of Pierce's data as showing symmetry is hard to understand for Pierce's symmetry was presumably assumed in the first place. His beautiful curves must have been fitted, for all he says of them, by eye, and when one attempts to treat his data statistically they do not check. *E. g.*, the numbers of cases for days 20 to 24 are respectively 495, 494, 502, 499, 498, and 497. Each should be 500. C. S. Pierce, *Report of U. S. Coast Survey*, 1870, 212ff.

⁵⁰ H. D. Williams, *Amer. J. Psychol.* 29, 1918, 219-226. See also Boring, *Psychol. Bull.* 15, 1918, 32f.

expressed the frequencies for recall of a material as a function of the number of repetitions of the material, but he felt, for certain cogent reasons that the student of memory will appreciate, that the repetition was not a satisfactory unit in terms of which to work. The psychophysical effect of an early repetition is presumably greater than that of a late one and Williams would have liked a more truly mnemonic unit. If he could have taken the normal curve *a priori* as the necessary function, he could have adjusted his results to it and have discovered a mnemonic unit that was some particular function of the number of repetitions. Such a relationship would have provided a knowledge of the effectiveness of successive repetitions and would have produced a method for use with other mnemonic measures. All measures eventually could have been interrelated, and a great advance made in the psychophysics of memory. Unfortunately, however, there was no justification for assuming the normal law, and, with both law and psychophysical unit unknown, Williams, as he realized, was placed in the dilemma of being able to reason from neither to the other. The way out, of course, was to give up the impossible, and to stick to the observed facts, leaving the frequencies to stand as a function of the number of repetitions.

Trabue, in attempting to establish a language-scale "with equal distances between steps," ran into the same difficulty.⁵¹ His material,—sentences to be completed by the Ebbinghaus method,—had no quantitative aspect except the number of arbitrarily chosen completion-tasks that could be successfully accomplished. He chose, therefore, to assume the normal law and to calibrate his material by reference to it, taking the probable error as the unit.⁵² His hint of a cogent reason back of the normal law stands quite unsupported.⁵³ Now the language-scales presumably do "work;" but if they do, it must be for the reason that this form of measurement, pretending to nothing more accurate than Quetelet achieved, counts his degree of rigor successful.

⁵¹ See foot-note above, p. 12.

⁵² Pp. 30ff. There is a further, somewhat obscured assumption, *viz.*, that the P. E. for one age equals that for another: see pp. 54ff. That the dispersion would not alter with age would hardly be expected; *cf.* J. B. Miner, *Deficiency and delinquency*, 1918, 252ff. At any rate the assumption is gratuitous, unless, as is always true, we argue that the unit is the premise in the case, and that premises are indisputable. The law of insufficient reason is usually applied to personal pre-dilection.

⁵³ A foot-note on p. 31 says of the normal law: "If this assumption is made the results that follow are in . . . close accord with known facts;" but there is nothing more.

These two papers are but examples of how the problem of the unit arises in very different fields of mental measurement and how their solution is balked by an inability to assume in advance some distributive function like the normal law.⁵⁴ In such cases the limitation is pretty clear, for we have to make an assumption and it takes but little argument to show that an assumption in itself can not be fact.

There is, however, a more subtle way in which the will to believe in the normal law gets in its work. Suppose we give up assumption and go to experience, in order to study a special group of related phenomena; and suppose we find that very many of these phenomena, as we measure them, are distributed every one in accordance with the normal law? May we not conclude by analogy that the same law applies to other phenomena? If a penny comes heads and tails equally often, will not a silver dollar? If stature follows the normal law, what about length of forearm?

The answer is that it all depends on how much one knows of the nature of the case. One may reason from the penny to the dollar, if one knows — inductively, from experience — that homogeneity and shape in the penny are the attributes that condition the law, whereas size, chemical constitution, color, *etc.*, are irrelevant. (Even so one may make a mistake; a magnetized steel disk in a magnetic field might give unexpected frequencies.) To reason from stature to length of forearm, however, requires that we know that the size of the forearm is conditioned in the same way as is the total length of the body. It is not enough to know that both fore-

⁵⁴ There is no need to assemble a list of studies where the error in question is committed. Often it is present only by implication. Pearson, for example, in his study of "The relationship of intelligence to size and shape of head, and to other physical and mental characters," *Biometrika* 5, 1906, 105-146, assumes the normal curve in order to provide a scale on which his results may be plotted. He is careful to point out that the assumption is merely for convenience and that it does not affect the chief business of the paper, *viz.*, the computation of correlation ratios (pp. 106f., 118). Very well. But then Pearson proceeds to set up a scale of units called "mentaces," with every unit 1/100 parts of a particular group that about equals the standard deviation. A complete scale of mentaces is suggested, ranging from -300 to +300. Offhand it looks like measurement, although the careful reader discovers that it is merely an illusion resulting from over-indulgence in the normal law.

Miner's discussion, *op. cit.*, 254-276, is here to the point. Miner sees the problem of the unit (255f.) and combats the Gaussian superstition (267ff., 272-276). His strictures on Pearson (272ff.) are tolerant. He does not seem, however, fully to appreciate (in spite of pp. 255 and 273f.) that the problem of the unit and of the form of distribution are one and the same.

arm and stature depend on the same factors; we must know that they depend upon them in the same way, that is to say, we must know that we are dealing with the same scale of growth. If the forearm were to vary as the square of x and stature as the cube of x , then prediction that the distribution of the one would be like the distribution of the other would be doomed to disappointment. With the coins, we know that the scales are comparable; each is expressed in percentage of heads. But in biological and psychological phenomena we can scarcely hope for any such certain analysis of the essential conditions.

Apparently then sometimes we may generalize and sometimes not. Can we make out a list of the classes of phenomena to which it is safe to say that the law applies? At least we may try.⁵⁵ And we shall have to consider five classes.

1. In the first place there are the *Glückspiele*,— coin tossing, card drawing, dice throwing, roulette, and the persistent urn or bag with the red and white balls inside, that introduces all students to the theory of probabilities. In such cases the presumption for the validity of the law is stronger than anywhere else. It seems that the red and white balls ordinarily do come out of the bag about equally often if they exist in equal quantities inside. Westergaard in 10,000 drawings got 5,011 white balls. Yet Pearson found that Weldon's 26,306 throws of dice exhibited "a bias toward the higher points." He calculated that the chances of 5 or 6 coming up on these dice exactly one-third of the time were less than two in 100,000, but that the chances of 5 or 6 coming up .3377 of the time

⁵⁵ What is needed is a comprehensive statistical study. If some one would get together only a hundred distributions based on large numbers (physical, biological, and mental distributions as found in nature, and distributions from such things as games of chance, tide tables, incommensurable numbers, *etc.*, as found in human civilization), and if he would treat them all accurately *by the same method*, applying some criterion of goodness of fit to the normal law (*cf.* Pearson, *Phil. Mag.*, *op. cit.*), we should begin to have some notion of the way in which all these things do behave. But the inexperienced statistician should not begin lightly on such a task. He must be prepared to use much judgment before results taken under very different conditions are comparable. The very decision as to the size of the class-interval in every case is tremendously baffling; and there are questions of weighting that are equally bothersome. When this task is done he ought to fractionate and do it all over, in order to see how the principle of the *Spielraum* as dependent on the law of large numbers works out. And after that, he has left the study of the effect of changes in the scale of units. How does the goodness of fit alter when the logarithm or the square root of the measure is substituted for the measure, and how much does it alter and under what conditions does it alter most? It is a large order and one that is well worth filling.

were as great as two in 16. "Oh well," we say, "you could not expect any real dice to be exactly homogeneous. There is not much difference between .3377 and $\frac{1}{3}$." No, and in any real experiment ideal conditions are never exactly realized. All experimental observation is only as accurate as the control of the conditions of observation.⁵⁶ The tendency with Pearson's method is to conclude from particular cases against a generality by a method more precise than the precision of the observed cases allows.⁵⁷ In general the *Glückspiele* seem to follow the law as well as the faulty conditions of their construction and operation permit.⁵⁸

2. It is generally maintained that the occurrence of digits in the expansion of an incommensurable number like π , e , $\sqrt{2}$, $\log 7$, etc. follows the normal law, and that all digits in the long run occur equally often. Here the law of insufficient reason is peculiarly tempting. If there is no relation at all between the intrinsic properties of a circle and the number of fingers on the two human hands which determined the decimal system, then why should not all digits turn up equally often in the expansion of π ? There is no reason why they *should* not. The *fact* is that in the first 707 digits of the decimal of π the sevens occur only three-fourths as often as they should, although the other digits do better by the law.⁵⁹ This case, however, is but a single one,⁶⁰ and, as Venn points out, the chances of one digit's deviating from equality as much as do

⁵⁶ Of course we *want* the die to come up equally often on every face. When it does not, we would rather blame the die than the law. But, though such a gaming 'instinct' may be built into human nature, it is nevertheless an insufficient reason for science. To blame the die because it does not fit the law is unfair; but to say that the die because of practical conditions does not disprove the law or even that it does approximately prove it, is allowable.

⁵⁷ But this is a paper in itself.

⁵⁸ For summaries of results, see Westergaard, *op. cit.*, 21-38; Pearson, *Chances of Death*, 1897, i, 11-25, 44-62. The failure of Monte Carlo to obey the law of chance seems not to apply to all roulette; *v. pp.* 57f. Westergaard's experiment is in *op. cit.*, 36f.; Pearson's computations are in *Phil. Mag.*, 5 ser., 50, 1900, 167ff. See also Fisher, *op. cit.*, 127-145.

⁵⁹ The actual frequencies, which Venn who discusses them (see note below) does not give, are: 0, 73; 1, 78; 2, 72; 3, 74; 4, 74; 5, 63; 6, 68; 7, 53; 8, 73; and 9, 79. The expansion of π is by W. Shanks, *Proc. Roy. Soc.* 21, 1873, 319.

⁶⁰ There is no reason why a division of a decade by a trinity should be prepotent of other trinities, that 10 divided by 3 should be 3.33333 . . . (in the decimal system; the duodecimal system gives a completed two-digit quotient); nor is there any reason why the product of 12345679 by 27 should be 333333333, *except* the absolutely adequate reason that such is the inherent nature of the case. But in the realm of probabilities these exceptions help to prove the rule.

the sevens is one in four.⁶¹ Edgeworth found that there was one chance in ten that the first 1,800 digits of the expansion of the fraction $1/1861$ were following the law of chance.⁶² Both he and Venn are convinced, however, that the rule of the coin is applicable to such series of digits, and indeed such is the presumption though the generality is here less well established than for games of chance.⁶³

3. In 1904, long after the publication of his first papers on skew variation in homogeneous material, Pearson made the following statement:

[Suppose] we assume a certain distribution of frequency for the character in human populations. This distribution of frequency is given by the Gauss-Laplacian normal curve of deviations from the mean. . . . Now the problem before us is the following one:—Is this assumption legitimate? It certainly is not true for organs and characters in *all* types of life. But it really does describe in a most remarkable manner the distribution of most characters in mankind. . . . I should be the last to assert that no human characters can be found that do not diverge sensibly from this Gaussian distribution. But I believe they are few, and that for practical purposes we may with nearly absolute safety assume it as a first approximation to the actual state of affairs.⁶⁴

Everything hinges here on the word “approximation.” The case is very different from that of the coin or of the incommensurable number. We do not know anything about the units with which we are working except that they are the units with which we are working. When we change from the penny to the dollar we keep the same scale, one of proportions of heads in a two-fold universe; when we change from π to $\sqrt{2}$ we also keep the same scale, one of proportions of digits in a ten-fold universe, but when we change from stature to the forearm we have no guarantee that the inch means the same thing biologically. And we must use a biological unit, if we are to bind our class together by a biological concept, if we are to predicate the normal law of all *human* characters. That the form of distribution is a function of the unit is unescapable. What we need to know, and what we do not know, is whether the differences that would occur with biologically reasonable

⁶¹ Venn, *op. cit.*, III-II8, 247f.

⁶² Edgeworth, *Cam. Philos. Trans.* 20, 1905, 128-130.

⁶³ See also S. Newcomb, “Note on the frequency of use of the different digits in natural numbers,” *Amer. J. Math.* 4, 1881, 39f.

⁶⁴ *Biometrika* 3, 1904, 142f. Many curves are given which *look* like good fits. The measures of goodness of fit obtained by Pearson and A. Lee, *ibid.* 2, 1903, 357-462, show great variation. There are equal odds that the statures of “mothers” follow the law, and only a few chances in a thousand that the forearms of “fathers,” “sons,” and “daughters” do. Pearson seems here, as we should like him to do elsewhere, to take his own criterion with a grain of salt.

variation of the biological unit from the physical unit are small enough to be included within the "approximation." A certain amount of alteration of the scale of measurement still leaves a sensibly normal curve. Have we any cogent grounds for assuming "with nearly absolute safety" that "the actual state of affairs" is not enough differently conditioned from those that we know about as to render it widely divergent from the normal law? This is a reasonable question, and within approximate limits it may be answerable. Certainly we may not generalize with Pearson on any other basis, and certainly Pearson neglects to provide this basis.⁶⁵

4. On such an insecure structure does Pearson attempt to build further. If the physical is normal, why not the mental?

I put the problem to myself as follows:—Assume the fundamental laws of distribution which we know to hold for the physical characters in man, and see whither they lead us when applied to the psychological characteristics. They must: (a) Give us totally discordant results. If so, we shall conclude that these laws have no application to the mental and moral attributes. Or, (b) Give us accordant results. If so, we may go a stage further and ask how these results compare with those for the inheritance of physical characters.⁶⁶

(The paper is on inheritance.) The assumption of the normal law gives results for the mental characteristics that look like the results for physical characters. Hence we may go ahead with our correlations in each case to indicate the degree of inheritance. The assumption accomplishes little beyond providing the scale for the printed graphs of intelligence; but that was in 1904.

In 1914 we find Pearson still arguing for a normal distribution of intelligence, and the occasions for his argument are certain empirical results which did not show a good agreement with the "Gaussian" law.⁶⁷ He is presenting the results of the application of the Jaederholm form of the Binet scale to school children of Stockholm. The samples are very small (261 normal and 301 defective children). There is one chance in 60 that the normal children give but a random deviation

⁶⁵ He is, in fact, not clear on the part played by the unit of measurement. He is concerned much over the normal law and but little over the arbitrary scale of measurement that determines in advance the distribution.

⁶⁶ *Biometrika* 3, 1904, 147. We have seen that Pearson also assumes the "Gaussian" curve for mental ability in *ibid.* 5, 1906, 105ff.

⁶⁷ Pearson and G. A. Jaederholm, *Mendelism and the problem of mental defect, II: On the continuity of mental defect*, 1914; Pearson, *Mendelism and the problem of mental defect, III: On the graduated character of mental defect, etc.*, 1914. These papers are viii and ix in *Questions of the day and fray*. On adjustment to the normal curve see II, 28-36, 46f.

from a properly "Gaussian" distribution; there is one chance in 20 that the normal children plus an ideal fraction of the defective children are properly distributed in the "Gaussian" manner. Eight-year old children taken alone, however, do show an even chance for the applicability of the "Gaussian" law. This is slim evidence, but it appears that it is not yet time to declare the Gaussian image a false god:

There is absolutely no reason why the Gaussian curve should be dogmatically asserted to apply to frequency-distributions of intelligence Still there are *a priori* suggestions that it should be tried. In the first place it does describe with a great degree of accuracy most physical measurements in man, and secondly, the Biometric school has found that it gives good results for many measures of intelligence. It is the view of the psychological joint-author of this paper that its comparative failure as applied to the present data lies rather in faults of the tests applied or in the method of applying them than in the non-Gaussian character of intelligence when adequately measured.⁶⁸

This quotation is as near a confession of what the Biometric School is trying to do as we are likely to get. It is frankly seeking always to see Gauss in nature. Its excuse in the case of mental phenomena for a prejudice in favor of this one of an infinity of possible distributions is that the Gaussian curve "does describe most physical measurements in man," and that "it gives good results for many measures of intelligence." The physical agreements, however, are approximate, as we have seen, and the facts with respect to intelligence are questioned.⁶⁹ The real difficulty comes, however, in the grounds for analogy. The implication is that we may reason from the physical to the mental because, in this case, both are "human," and that we may reason from one test to another because both are tests "of intelligence." Now we have no distributions of 'humanness,' but only of measurements of human characters; and we have no distributions of intelligence, but only of particular measurements of intelligence.⁷⁰ Moreover, in every case the measure is arbitrarily chosen; and it is not even strict-

⁶⁸ *Opp. cit.*, *Mendelism, etc.*, II, 46.

⁶⁹ Miner, *op. cit.*, 267-279, reviews the evidence for a normal distribution of tested intelligence, and concludes: "In spite of these arguments and the evidence of asymmetry of measurements at least of some periods of life it is to be noted that current opinion is probably contrary to this hypothesis [hypothesis, the normal law], although, as I believe, because it has been mainly concerned with those who are not of extreme ability:" p. 275.

⁷⁰ Cf. here Miner's distinction between "physical units" and "more equivalent biological units," *op. cit.*, 255. We have noted that Miner does not take advantage of the full force of the distinction; the physical unit may be irrelevant to the problem in hand.

ly a measure, in the sense of being the sum of equal units, unless we make it so on *a priori* grounds. In an intelligence-test, for example, we have no notion whether the difference between 99 and 100 points is the same difference in intelligence as the difference between 9 and 10 points, unless we are ready to define intelligence arbitrarily as the ability to achieve points in that particular test.⁷¹ Now if we define intelligence with respect to one test, we can not assert *a priori* that any other test is a measure of intelligence, and should therefore show the same distribution. *A posteriori* we may show by correlation that a second test measures an ability similar to intelligence as defined by the first test, but in such a process we demonstrate empirically the similarity of distribution. There is no valid reason for expecting normality in one intelligence-test because it has been found in another; and the same reasoning applies to the argument from the physical to the mental.⁷²

We are very far, then from a general conclusion that intelligence, a mental capacity comprehensible apart from the particular instrument by which it is tested, is normally distributed. Still farther are we from stating that mental abilities, whatever their nature, follow the normal law.

5. We have to investigate one more class of phenomena, the psychophysical judgments. In this case no claim for generality has been presented, but the trend of the facts bespeaks attention. Urban found that the phi-function of gamma is a good hypothesis for the psychometric function in lifted

⁷¹ No one who has ever "made up" an intelligence test ought to have any difficulty in appreciating the arbitrariness of the calibration. The tasks of the test are in the first place sheer casual invention. They are put into a scale and tried out. Alterations are made to "improve" the frequencies of success in different component parts, but no one can work accurately from the frequencies to the scale without presupposing the form of distribution. There is never any attention paid to making equal increments of intelligence measured by equal increments of the scale, for the excellent reason that no one has any such definite quantitative conception of intelligence. To call a year in mental age a unit is a fallacy of this sort. See Pearson, *Mendelism, etc.*, II, 36ff.; Miner, *op. cit.*, 260ff. There is no evidence (nor are the physical growth curves in any way relevant) that annual increments of intelligence are equal in any other sense except in that they are annual. Of course a year is a year, if "annual-ness" is what we are after.

⁷² Pearson has appointed himself "watch-dog" of "logic in scientific procedure", *Mendelism, etc.*, III, 3f.; but he seems to need to make friends with one sophistical wolf ("an appearance of knowledge where we are as yet in a state of ignorance") in order to attack another.

weights and in certain acoumetric experiments.⁷³ Now the phi-function of gamma is simply the integral form of the normal curve. To say that the phi-function of gamma is the psychometric function is equivalent to saying that the dispositional variations of the psychophysical organism, when measured in a scale of units proportional to the scale in which the stimulus is measured, follow the normal law. If the lifted weights are measured in grams and the psychometric function approximates the phi-function of gamma, then the organism varies in the amount of its disposition for judging "heavier" or "lighter" according to normal law, provided always that the amount of this disposition is measured in grams. This relationship is scarcely obvious and the reader may need to read elsewhere,⁷⁴ before he can accept it. The essential thing to understand is that the applicability of the phi-function of gamma means that the organism is varying somehow in accordance with the normal law, but that the normality of its variation is as much a function of the unit in which the variation is measured as it is in any of the other cases which we have considered.

Now can we generalize? Can we ever say that all psychophysical judgments tend to obey the normal law? Urban makes no such attempt; rather is he against theorising and for description. "The nature of the dependence," he writes, "is not known and cannot possibly be deduced by any considerations *a priori*,"⁷⁵ and he refers us for the function to "the results of observation." This, to the scientist, is a refreshing return to facts from the realm of logic and mathematics; but is not final. Something can be said for generalization.

Unfortunately for ease of exposition, there are two kinds of psychophysics. In the one kind stimuli are presented and the observer reports upon the sense-impressions. In the case of the stimulus-limen, he simply reports sensation or its absence. He is making a brief introspection. In the case of the differential limen he makes a judgment upon the relative degrees of the two sensations, but the judgment is actually little more than a comparative description; the ob-

⁷³ On lifted weights: Urban, *op. cit.*, 257ff.; *Arch. f. d. ges. Psychol.* 16, 1909, 168-227, espec. 224f. On acoumetry, H. Keller's experiments, *Psychol. Stud.* 3, 1907, 49-89; and Urban's discussion of them, *Arch. f. d. ges. Psychol.* 18, 1910, 400-410. Keller used Wundt's fall phonometer, and the all-important unit of stimulus, which is only just mentioned, is the cm. The psychometric functions of the lifted weights are for grams.

⁷⁴ Boring, *Amer. J. Psychol.* 28, 1917, 465ff.

⁷⁵ *Psychol. Rev.* 17, 247.

server may be said to be introspecting here too.⁷⁶ In the other kind of psychophysics the observer's attention is directed upon the stimulus and not upon his own mental processes. In no sense does he introspect. Always he is making judgments of the stimulus. In introspective psychology this attitude is called the "stimulus error," but in psychophysics it may be in place. Urban's experiments are of this kind,⁷⁷ as has been much other psychophysical work. There is reason to believe, however, that psychology will profit more in the long run by work performed under the first attitude described.

In the case of the "introspective" psychophysics we are not to expect a generalization, because everything depends on the unit. Were the psychometric function to approximate the phi-function of gamma ten times, there would be no reason to expect it the eleventh. Suppose the eleventh case were the differential limen for auditory intensity with the Fechner sound pendulum as the instrument. Will the normal law hold? Perhaps when we use as the unit $\sin^2 (\Theta/2)$. If so, it will not hold if we use Θ as the unit. If it holds for neither, it may hold for some other function of Θ . The reports upon sensations depend upon Θ and the momentary disposition of the organism, but in no way upon the scale of units chosen. It is mere luck if the normal law applies. But whatever luck we get is good luck, for the psychometric function gives us the law of the dependence of dispositional variation upon the scale of measurement which we have chosen. A little knowledge of sensory mechanisms is worth the loss of a generality.

When the psychophysical judgments are directed upon the

⁷⁶ Cf. the nature of the constant psychophysical attitude set up in S. S. George's experiment, *Amer. J. Psychol.* 28, 1917, 1-37.

⁷⁷ In the discussion of Ueber einige Begriffe und Aufgaben der Psychophysik, *Arch. f. d. ges. Psychol.* 30, 1913, 113-152 (cf. also, e. g., *Amer. J. Psychol.* 24, 1913, 274), Urban would appear to maintain that psychophysics deals with describable mental processes; but the implications in the accounts of his experiments, the lack of a reproduction of definite instructions given the observers, and the failure to discuss the introspective competence of the observers, all indicate that the observers were, or at least may have been, making judgments of stimulus rather than of mental process. See *The Application of Statistical Methods to the Problems of Psychophysics*, 1908, 5, 14; *Arch. f. d. ges. Psychol.* 15, 1909, 264f. In *Psychol. Rev.* 17, 229, he speaks of "the judgments of a subject who compares two stimuli." And indeed in the systematic schema of *Arch.* 30 (*op. cit.*) he takes account (1) of mental contents and processes, (2) of physiological conditions and processes, and (3) of physical conditions and processes, but leaves no place for meanings, awarenesses, the stimulus in relation to the subject. It is doubtful if Urban sought at all to distinguish between *Beschreibung* and *Kundgabe* in the psychical aspect of the psychophysical system.

stimulus it may be that we can generalize, for now the scale and the judgment are related. With the sound pendulum we might be asked to judge the height of the fall, which is measured by $\sin^2 (\theta/2)$, or we might be asked to judge the angle. In the two cases we would be judging entirely different things, and it is quite possible that the normal law might apply in each. If we found that it applied in a large number of cases, then we might generalize by saying that errors in the judgment of a stimulus, that is to say, errors of observation, are distributed normally. This conclusion, indeed, is Urban's: "We thus obtain the remarkable result that the foundations of the theory of the errors of observation are found in the theory of psychophysical measurement."⁷⁸ In other words we would demonstrate psychophysically Gauss' original assumption that positive and negative errors are equally likely.

If there is any general answer to the question: When can we generalize?—it is this. We may expect the normal law to hold within an entire class of phenomena, when it holds for a number of cases within the class, provided always that all variations of members of the class are capable of being expressed as variations of some common denominator. In other words, we must be able to see the same thing varying in different situations. With the penny and the dollar, the law holds for the turning up of heads. The things that make the penny different from the dollar, we know to be irrelevant. In lifted weights and acoumetry we may know, by a control of attitude, that we are studying, not kinaesthesia and auditory sensation, but errors of observation. If Urban did know this, then he was supplied with the only possible ground for expecting the one result on the basis of the other.

4. THE LOGICAL POSSIBILITIES (Conclusion)

At last we are ready to take account of stock! There are four logical possibilities, but the circumstances of the case are such that it appears unlikely that the actual mental measurement can make much use of more than one. Let us see.

1. In the first place, there is the possibility of determining *a priori* the form of distribution. This course, applied to the normal law, is popular, but it is usually without sanction. In quantitative psychology we should resist it. We may not assume the normal law in an interpretative study except upon

⁷⁸ *Psychol. Rev.* 17, 243; but Urban is being betrayed into generalizing. His statement is based only on the lifted weights and perhaps the acoumetric experiments. The validity of generalization under this attitude is hardly to be demonstrated so easily.

valid grounds. We have seen that the law of insufficient reason furnishes no such grounds. Knowledge simply does not come out of ignorance. The only scientific grounds that are presumptive of the normal law are an intimate knowledge of the constitution of the particular case and of the function that frequencies of occurrence are of that constitution. Such a function comes in science from experience, and, whenever it is thoroughly understood, from experiment. On such cogent grounds we may conclude *a priori* to the normal law or some other form of distribution, and we are following such a logical process in the few cases where we find it legitimate to generalize about the form of distribution for a given class of phenomena.

Two corollaries follow.

(a) If we knew that the normal law had to hold in a particular case, then we could experiment with an arbitrarily chosen unit and determine, by working from the law, the function that the "true" psychological unit, which is following the law, is of the arbitrary unit of the experiment. Any such hope, however, is bound to be illusory, for a knowledge of the constitution of the case involves *a priori* the knowledge of this unit. We may say that, if the dollar is like the penny, we shall get 50% heads; but we can not reverse the argument and say that, because we get 50% heads, the dollar is like the penny.⁷⁹

(b) There is at present very little prospect in the field of mental measurement that our knowledge of the psychological constitution of a mental function or process will be sufficient to enable us to begin work upon a problem with the assurance that the resulting distributions must be normal. There is a bare hint of an exception, that awaits verification, in the psychometric functions of errors of observation. But the evidence indicates that instances, where there are cogent reasons for assuming the normal law in advance of empirical determination, will be extremely rare for a long time to come.

2. We may begin in mental measurement with the psychological unit, for if we can not determine the unit from the distribution, we may, nevertheless, determine the distribution from the unit. Such, in fact, is the necessary scientific order.

⁷⁹ *I. e.*, in physical constitution. We are not denying inverse probability, but insisting merely that a statement of physical constitution is more complex than a statement of frequency and that the same frequency may be variously conditioned. If we know only that we get 50% heads, we can not reason that the underlying physical constitution is that of the dollar or the penny. For all we can tell we may be dealing with an asymmetrical loaded die, with one head and five tails, that belongs with the penny only in the matter of frequency of heads and not in general physical constitution.

The great difficulty is, as we have just pointed out, to find anything that we may properly call a psychological unit. The sense-distance is such a unit. It is a unit of measurement that is mental *per se*.⁸⁰ Most of the sensory work in psychology, however, has had to do with the determination of limens and has not made use of the mental unit at hand. The extensive psychophysics of memory has not achieved a psychological unit, nor have the other departments of the psychology of process. In the psychology of capacity or function the case is worse. For example, with intelligence, which is the mental capacity most often measured, we have seen that, not only is there no attempt to make equal increments on an intelligence scale correspond to equal increments of intelligence, but that the concept of intelligence is so vague that any such accurate quantitative relationship is in practice almost meaningless. We are not, however, at our rope's end.

3. If a psychological unit is not to be achieved we may use a "physical" unit, that is to say, some arbitrarily chosen aspect of an arbitrary scale of measurement. Such units are the year (mental age), the second (mental tests where time is the measure), the item or task (where the number of points is the number of unit tasks completed, *e. g.*, the U. S. Army intelligence tests), the gram (lifted weights), the syllable (memory experiments), and so on. Every such unit is arbitrary. There is no evidence that equal increments of its scale correspond to equal increments of the psychological entity measured. When we define intelligence, in order to gain definiteness of conception, as ability in some particular test, we are simply substituting a "physical" unit for the ideal, but unachievable, psychological unit. Now the application of a unit that is not psychological to a quantity that is psychological does not yield a measure of the quantity. It will place the quantity in a given position upon the arbitrary scale, and will determine the rank-orders of a number of quantities so placed, but the assumption that rank-orders tell the amount of a given quantity, or, if the zero be unknown, of a given increment, is, of course, unwarranted.

There is one case, of frequent occurrence in the behaviorism

⁸⁰ Cf. E. B. Titchener, *Experimental psychology, Quantitative student's manual*, 1905, xix-xxxvii. The thing is so simple when it is seen! In physical science the nature of the unit is given in the pre-suppositions of the science, and in psychology it is given mentally in the sense-distance. It is the indefiniteness of the mental function that makes trouble and gives rise to the attempt to create a unit in retrospect by an appeal to the hypothesis of the normal law. Since this paper deals with the appeal to the normal law, we may, however, pay homage to the sense-distance and proceed.

of mental tests, where the "physical" and psychological units seem to become identical. Suppose we are studying the learning of typewriting and express our results in the number of words written per minute. Do not words-per-minute truly measure ability in typewriting? Here we are involved in a question of point of view. Words-per-minute measure the product of typewriting. They may perhaps be said to measure ability in typewriting from the point of view of the employer of a stenographic force; they are the ability of his office force over against a given job. A behavioristic psychology that identified behavior with mere product, without any reference at all to the conditions of the product, might take the point of view of the employer; but such a psychology would achieve only a physics, or more likely a common sense, of typewriting. Any definition of behavior as *response to a situation*⁸¹ brings in the behaving organism and changes the ground. We can not say that words-per-minute truly measure the behavior of the psychophysical organism or its response to the situation. Does an increment of ten words per minute mean the same change in ability or in behavior or in response when added to an ability of 50 words per minute as when added to an ability of 100 words per minute? To the employer, yes; but not to the psychologist, or, presumably, to the organism. It is indeed an heroic measure, when we can not make the unit psychological, to attempt to a-psychologize psychology.

4. We are left then with the rank-orders of our psychological quantities, given by reference to a fixed but arbitrary extra-psychological scale; and it is with these rank-orders that we must deal. We are not yet ready for much psychological measurement in the strict sense. It is true that psychophysics may yield a unit and also true that ordinarily it does not. The sense-distance has had only a limited use, and we do not yet know that the stimulus-distance (for it is upon stimulus-distances that the psychometric functions of errors of observation are founded) constitutes a true unit in a behavioristic psychophysics. The case, however, is not so bad as it seems at first. There is nothing new in the contention that mental measurement is impossible,⁸² whereas now we do gain the assurance that rank-orders at least are validly demonstrable. And there

⁸¹ Cf. the point of view of behavior presented by E. B. Holt, *The Freudian wish*, 1916, espec. 153ff. Behavior is *what* the individual is *doing*, not that he is moving. We can perhaps measure movement, but we have yet to conceive the unit for "doings."

⁸² Indeed, in confining ourselves to a consideration of the part played by the normal law, we are leaving a famous controversy untouched. Cf. O. Klemm, *A history of psychology*, trans. 1914, 150-155, 232-267.

is a great deal that can be done with rank-orders. We can deal with frequencies, medians, and quartiles. For example, it is considerable to know of two groups that the lower quartile of the ranks in one overlaps and is practically coincident with the upper quartile of the other. What we must remember, however, is that we are dealing with the statistics of medians, quartiles, contingencies, and correlation ratios; not with the statistics of averages, standard deviations, coefficients of correlation, and linear regressions. All those statistical constants, that imply a scale of equivalent units, violate in use the conditions of the case and lead to a precision of result that is an artifact. The serial constants, that do not presuppose a unit, yield less intricate resultants, but they present a rougher picture that represents truly the rough material which they describe.⁸³

The initial error in the application of the theory of probabilities was the assumption of the law of insufficient reason. It was wrongly supposed that knowledge could somehow be wrought out of ignorance. This very error, however, has never been routed. It has gone on, multiplying mischief. The substitute for insufficient reason is cogent reason. The more we know of the intimate nature of the entity with which we are dealing the more accurate and complete can our descriptions become. But, if in psychology we must deal—and it seems we must—with abilities, capacities, dispositions and tendencies, the nature of which we can not accurately define, then it is senseless to seek in the logical process of mathematical elaboration a psychologically significant precision that was not present in the psychological setting of the problem. Just as ignorance will not breed knowledge, so inaccuracy of definition will never yield precision of result.

⁸³ Cf. C. J. West, *Introduction to mathematical statistics*, 1918, 94.